



```

AC      P09801.1 (Rel. 10, created)
DT      01-MAR-1989 (Rel. 10, last sequence update)
DT      01-MAR-1989 (Rel. 10, last annotation update)
DE      15-JUL-1999 (Rel. 38, last annotation update)
DR      VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).
OS      Gossypium hirsutum (Upland cotton).
OC      Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC      Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC      Malvales; Malvaceae; Gossypium.
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;
RT      "Developmental biochemistry of cottonseed embryogenesis and
RT      germination. xviii. cdna and amino acid sequences of the members of
RT      the storage protein families."
RL      Plant Mol. Biol. 7:475-489(1986).
CC      -I- FUNCTION: SEED STORAGE PROTEIN.
CC      -I- SUBCELLULAR LOCATION: CYTOSOL; VACUOLAR PROTEIN BODIES.
CC      -I- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC      CONVICILIN, CONGLYCININ, ETC.).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/ib-fib.ch).
CC      or send an email to license@ib-fib.ch).
CC      -----
DR      EMBL: M16891: AAA33071.1; -.
DR      PIR: A30838; FWCNAB.
DR      HSSP: P50477; ICAX.
DR      INTERPRO: IPR001113; -.
DR      Pfam: PF00546; Seedstore_7s; 1.
KW      Seed storage protein; Signal.
FT      CHAIN          1          25
FT      SIGNAL         1          25
FT      CHAIN          26          588      VICILIN C72.
SO      SEQUENCE      588 AA;  69729 MW;  63E699B23ABADEE CRC64;

```

```

Query Match Similarity      38.6%; Score 132; DB 1; Length 588;
Best Local Similarity       35.8%; Pred. No. 6,7e-06;
Matches    29; Conservative 11; Mismatches 19; Indels 22; Gaps 3.

OY      2   ORQYOCCGCGCOEOOGGOOGEOOOOCCKCKEYOKEO-----ERGEHNNYHN 46
          |||::||| | :| :| :||| | :| :| :|
Db      121  ORQFECGCCHCCHCCEORRPRKQQCVCEGRERYQENPWRREEREAEAEETEEGEOEOSHN 180
          |||::||| | :| :| :||| | :| :| :|

OY      47   ---HKK---NRSEEGGOR 60
          ||:: | || | |
Db      181  PPHFHRRSPQSFRREHGNGFR 201

RESULT      3
GBF_DICDI ID GBF_DICDI STANDARD: PRG: 708 AA.
AC P36417;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE G-BOX BINDING FACTOR (GBF).
GN GBFA.
OS Dictyostelium discoideum (Slime mold).
CC Eukaryota; Dictyostelidia; Dictyostelium.
(1)
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-AX3;
RX MEDLINE=94170994; PubMed=812561;
RA Schnitzler G.R., Fischer W.H., Firtel R.A.;
RT "Cloning and characterization of the G-box binding factor, an
late development component in Dictyostelium.";
```

```

RL   Genes Dev 8:502-514 (1994).
CC   -i- FUNCTION: CAMP-RESPONSIVE TRANSCRIPTIONAL ACTIVATOR REGULATING
CC   LATE GENE EXPRESSION. ESSENTIAL COMPONENT OF THE DEVELOPMENTAL
CC   SWITCH BETWEEN EARLY AND LATE DEVELOPMENT. BINDS TO A NUMBER OF
CC   C/CAAT-RICH GENE REGULATORY ELEMENTS.
CC   -i- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@sib-sib.ch).
CC   -----
CC   C
CC   CC
CC   DR   EMBL; L29075; AAA21021.1; -.
CC   DR   TRANSFAC; T00315; -.
CC   DR   DICTYDB; DD02046; GBFA.
KW   Transcription regulation; Activator; DNA-binding; Nuclear protein;
KW   Repeat.
KW   REPEAT 339 368 1.
ET   REPEAT 481 510 2.
ET   DOMAIN 1 21 POLY-SER.
ET   DOMAIN 115 263 GLN-RICH.
ET   DOMAIN 270 292 POLY-ASN.
ET   DOMAIN 549 557 POLY-ASN.
SQ   SEQUENCE 708 AA; 79268 MW; BAB6DE04EFAACCA CEC64;

```

```

Query Match          27.2% Score 93; DB 1; Length 708;
Best Local Similarity 31.0%; Pred. No. 0.04;
Matches 18; Conservative 14; Mismatches 26; Indels 0; Gaps 0;

OY      2   OROVOCCGRCQEDQQGGQRERDQCCKRKSEDEYKKEEGENHENYHNKKRNEEESGOQ 59
         I : I I I : I I I : I I I : I : I : I : I : I : I : I : I : I : I
Db       190 OHNHOOOHHHOOOHNOHNNHOHHNDOOOHHNHOHHNHHNSHPROOHNNHHNQDHQOQ 247

RESULT    4
INVO_CANFA INVO_CANFA STANDARD: PRT; 265 AA.

AC        PI8174;
DT        01-NOV-1990 (Rel. 16, Created)
DT        01-NOV-1990 (Rel. 16, Last sequence update)
DT        01-FEB-1996 (Rel. 33, Last annotation update)
DE        INVOLUCRIN.
GN        IVL.
OS        Canis familiaris (dog).
OC        Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN        [1]
RP        SEQUENCE FROM N.A.
RX        MEDLINE=90348475; PubMed=2385171;
RA        Tseng H., Green H.;
RT        "The involucrin genes of pig and dog: comparison of their segments of
            repeats with those of prosimians and higher primates.";
RL        Mol. Biol. Evol. 7:293-302(1990).
SC        -I FUNCTION: INVOLUCRIN IS A KERATINOCTYE PROTEIN THAT FIRST APPEARS
            IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO
            MEMBRANE PROTEINS BY TRANSGLUAMINASE. ALL THAT RESULTS IN THE
            FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.
CC        -I TISSUE SPECIFICITY: PRESENT IN KERATINOCTYES OF EPIDERMIS AND
            OTHER STRATIFIED SQUAMOUS EPITHELIA.
CC        -----
CC        THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC        between the Swiss Institute of Bioinformatics and the EMBL outpost.
CC        The European Bioinformatics Institute. There are no restrictions on its
CC        use by non-profit institutions as long as its content is in no way
CC        modified and this statement is not removed. Usage by and for commercial
CC        entities requires a license agreement (See http://www.isb.ch/announce/
            or send an email to license@isb.sib.ch).
CC        -----
EMBL: M34442::AAA30853.1; .

```



```

CC      -I- SIMILARITY: BELONGS TO THE SNF5 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its use by
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      DR   EMBL; M36482; AAA35062.1; -.
CC      DR   EMBL; X76053; CAAS3652.1; -.
CC      DR   EMBL; Z36158; CAAB8254.1; -.
CC      DR   PIR; S44551; RGRBYS.
CC      DR   PIR; S39145; S39145.
CC      DR   SGD; S0000493; SNF5.
CC      KW   Transcription regulation; Activator; Nuclear protein.
CC      FT   DOMAIN 31 270 GLN-RICH.
CC      FT   DOMAIN 72 132 PRO-RICH.
CC      FT   DOMAIN 272 324 PRO-RICH.
CC      FT   DOMAIN 489 588 ASP/GLU-RICH (ACIDIC).
CC      FT   DOMAIN 714 882 PRO-RICH.
CC      FT   DOMAIN 755 798 ARG/LYS-RICH (BASIC).
CC      FT   CONFLICT 564 564 E -> D (IN REF. 1).
CC      SO   SEQUENCE 905 AA: 102557. MW: 428754a648bd1a5 CRC64;

```

Query Match	26.3%	Score 90:	DB 1:	Length 905;
Best Local Similarity	29.1%	Pred. No.	0.096;	
Matches 16;	Conservative 20;	Mismatches 19;	Indels 0;	Gaps 0;

QY 4 OYOCOCGCGEOOGGORGROOORCKCMECYKCEERGENHENHNHKKKNSEEEGO 58  
| | | | : | | | : | | | : | : | : : : : :  
Db 216 QIQOOQQOQQOQQHQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 270

RESULT	7
U2R2_HUMAN	
ID	U2R2_HUMAN
STANDARD;	
PRT;	482 AA.

```

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE U2 SMALL NUCLEAR RIBONUCLEOPROTEIN AUXILIARY FACTOR 35 KDA SUBUNIT
DE RELATED-PROTEIN 2.
GN UZAF1-RS2 OR UZAF1RS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN.
RX MEDLINE=96163878; PubMed=8586425;
RA Kitagawa K., Wang X., Hatada I., Yamaoka T., Nojima H.,
RA Inazawa T., Abe T., Mitsuya K., Oshimura M., Monden M.,
RA Mukai T.;
RT "Isolation and mapping of human homologues of an imprinted mouse gene
RT Uzafl-rs1."
RL Genomics 30:257-263(1995).
CC -1 SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1 SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -1 SIMILARITY: TO MAMMALIAN SPLICING FACTOR UZAF 35 KDA SUBUNIT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D49677.1; BAA08533.1; -.
DR MIM; 300028; -.

```

DR	INTERPRO: IPR000504; -	
DR	INTERPRO: IPR000571; -	
DR	PFAM: PF000076; rrm; 1	
DR	PFAM: PF00642; zf-CCH; 1	
DR	PROSITE: PS50102; RRM; 1.	
KM	Nuclear protein; Ribonucleoprotein.	
FT	DOMAIN	46
FT	DOMAIN	118
FT	DOMAIN	123
FT	DOMAIN	482 AA;
SEQUENCE	1DACC8ACCA4727266	CRC64;

Query Match	25.4%	Score 87	DB 1	Length 482
Best Local Similarity	25.4%	Pred. No. 0	11	
Matches 15	Conservative 22	Mismatches 22	Indels 0	Gaps 0

[illegible]

RESULT		8
WCI_NEUCR	ID	WCI_NEUCR
STANDARD:		PRT; 1154 AA.

DT 01-NOV-1997 (Rel. 35, last sequence update)  
DT 30-MAY-2000 (Rel. 39, last annotation update)  
DE WHITE COLLAR 1 PROTEIN (WCI).

OS Neurospora crassa.  
OC Fungi; Ascomycota; Sordariales; Sordariaceae; Neurospora.  
RN Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae; Neurospora.  
[1]

RA MEDLINE=96203083; PubMed=6612589;  
RX Ballarín P., Vitorioso P., Magrelli A., Talora C., Cabibbo A.,  
RA Macino G.;

EMBO J. 15:1650-1657(1996)  
-1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR INVOLVED IN LIGHT  
REGULATION. BINDS AND AFFECTS BLUE LIGHT REGULATION OF THE AL-3

```
CC ACTIVATE TRANSCRIPTION
CC - SUBUNIT: HETERODIMER OF WC1 AND WC2 (POTENTIAL).
CC - SUBCELLULAR LOCATION: NUCLEAR.
```

CC GENE EXPRESSION.

CC -1 SIMILARITY: BELONGS TO THE GATA-TYPE TRANSCRIPTION FACTOR FAMILY.

CC -1 SIMILARITY: CONTAINS A PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.

between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC send an email to [license@lsb.slb.ch](mailto:license@lsb.slb.ch).  
 OR EMBL; X94300; CAA63964.1; -.

```
DR INTERPRO, IPR000679, -.
DR INTERPRO, IPR001610, -.
PEAM: PF00320; GATA: 1.
```

DR PROSITE; PS00344; 1. GATA\_ZN\_FINGER\_1; 1. DR PROSITE; PS50114; 1. GATA\_ZN\_FINGER\_2; 1. DR Transcription regulation; Activator; DNA-binding; Zinc-finger;

KW Nuclear protein; Repeat.











RN [14]  
 RP VARIANT LNCAP ALA-877.  
 RX MEDLINE=91083633; PubMed=2260966;  
 RA Veldschoote J., Ris-Stalpers C., Kuiper G.G., Jenster G.,  
 RA Berrevoets C., Claassen E., van Rooij H.C.J., Trapman J.,  
 RA Brinkman A.O., Mulder E.; "A mutation in the ligand binding domain of the androgen receptor of  
 RT human LNCAP cells affects steroid binding characteristics and  
 RT response to anti-androgens."  
 RL Blochem. Biophys. Res. Commun. 173:534-540(1990).  
 RN [15]  
 RP VARIANT CAIS MET-866.  
 RX MEDLINE=91186983; PubMed=2082179;  
 RA Brown T.R., Lubahn D.B., Wilson E.M., French F.S., Migeon C.J.,  
 RA Corfen J.L.; "Functional characterization of naturally occurring mutant androgen  
 RT receptors from subjects with complete androgen insensitivity."  
 RL Mol. Endocrinol. 4:1759-1772(1990).  
 RN [16]  
 RP VARIANT CYS-774.  
 RX MEDLINE=91310758; PubMed=1856263;  
 RA Marcelini M., Tilley W.D., Zoppi S., Griffin J.E., Wilson J.D.,  
 RA McPhaul M.J.; "Androgen resistance associated with a mutation of the androgen  
 RT receptor at amino acid 772 (Arg->Cys) results from a combination of  
 RT decreased messenger ribonucleic acid levels and impairment of  
 RT receptor function."  
 RL J. Clin. Endocrinol. Metab. 73:318-325(1991).  
 RN [17]  
 RP VARIANT CAIS ASN-695 AND HIS-695, AND SEQUENCE OF 629-723 FROM N.A.  
 RX MEDLINE=92131007; PubMed=1775137;  
 RA Ris-Stalpers C., Trifiro M.A., Kuiper G.G., Jenster G., Romalo G.,  
 RA Sai T., van Rooij H.C., Kaufman M., Rosenfield R.L., Liao S.;  
 RT "Substitution of aspartic acid-686 by histidine or asparagine in the  
 RT human androgen receptor leads to a functionally inactive protein with  
 RT altered hormone-binding characteristics."  
 RL Mol. Endocrinol. 5:1562-1569(1991).  
 RN [18]  
 RP VARIANTS CAIS AND PAIS.  
 RX MEDLINE=9338440; PubMed=1307250;  
 RA Batch J.A., Williams D.M., Davies H.R., Brown B.D., Evans B.A.J.,  
 RA Hughes I.A., Patterson M.N.; "Androgen receptor gene mutations identified by SSCP in fourteen  
 RT subjects with androgen insensitivity syndrome."  
 RL Hum. Mol. Genet. 1:497-503(1992).  
 RN [19]  
 RP VARIANT CAIS VAL-787.  
 RX MEDLINE=92235226; PubMed=1569163;  
 RA Nakao R., Hagi M., Yanase T., Ogo A., Takayanagi R., Katsube T.,  
 RA Fukumaki Y., Nawata H.; "A single amino acid substitution (Met-786->Val) in the steroid-  
 RT binding domain of human androgen receptor leads to complete androgen  
 RT insensitivity syndrome."  
 RL J. Clin. Endocrinol. Metab. 74:1152-1157(1992).  
 RN [20]  
 RP VARIANT LNCAP ALA-877.  
 RX MEDLINE=92222955; PubMed=1562539;  
 RA Veldschoote J., Berrevoets C.A., Ris-Stalpers C., Kuiper G.G.,  
 RA Jenster G., Trapman J., Brinkman A.O., Mulder E.;  
 RT "The androgen receptor in LNCAP cells contains a mutation in the  
 RT ligand binding domain which affects steroid binding characteristics  
 RT and response to antiandrogens."  
 RL J. Steroid Biochem. Mol. Biol. 41:665-669(1992).  
 RN [21]  
 RP VARIANT MET-730.  
 RX MEDLINE=92335289; PubMed=1631125;  
 RA Newmark J.R., Hardy D.O., Tonb D.C., Carter B.S., Epstein J.L.,  
 RA Isaacs W.B., Brown T.R., Barrack E.R.;  
 RT "Androgen receptor gene mutations in human prostate cancer."  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:6319-6323(1992).  
 RN [22]  
 RP VARIANT CAIS VAL-754.  
 RX MEDLINE=93372806; PubMed=8103398;

RA Lobaccaro J.-M., Lombroso S., Ktari R., Dumas R., Sultan C.;  
 RT "An exonic point mutation creates a Maelil site in the androgen  
 RT receptor gene of a family with complete androgen insensitivity  
 RT syndrome."  
 RL Hum. Mol. Genet. 2:1041-1043(1993).  
 RN [23]  
 RP VARIANT CAIS ARG-807.  
 RX MEDLINE=94108430; PubMed=8281140;  
 RA Adeyemo O., Kallio P.J., Palvimo J.J., Kontula K., Jaane O.A.;  
 RT "A single-base substitution in exon 6 of the androgen receptor gene  
 RT causing complete androgen insensitivity: the mutated receptor fails  
 RT to transactivate but binds to DNA in vitro."  
 RL Hum. Mol. Genet. 2:1809-1812(1993).  
 RN [24]  
 RP VARIANT PAIS VAL-743.  
 RX MEDLINE=93315568; PubMed=8325932;  
 RA Nakao R., Yanase T., Sakai Y., Hagi M., Nawata H.;  
 RT "A single amino acid substitution (Gly743 -> Val) in the steroid-  
 RT binding domain of the human androgen receptor leads to Reifenstein

Query Match 24.4%; Score 83.5; DB 1; Length 919;  
 Best Local Similarity 32.3%; Pred. NO. 0.4; Mismatches 17; Indels 11; Caps 2;  
 Matches 21; Conservative 17; Mismatches 16; Indels 11; Caps 2;  
 QY 1 LORQYQCGRCQEQOQGGREQOQCKRCWEQYKEDRGCHENYHNKKNRS-----EE 54  
 Db 57 LQQQOQO-----QQQOQOQOQOQOQOQFSPROOQOQOQGDGSPQAHRRPCTGLVLDDE 111  
 QY 55 EECQO 59  
 Db 112 QQPSQ 116

Search completed: March 1, 2001, 16:16:54  
 Job time: 434 sec

